

NÁSKOK  
DÍKY  
ZNALOSTEM

PROFINIT

B0M33BDT

Technologie pro velká data

Supercvičení – Hive

Sergej Stamenov, Jan Hučín

30. 10. 2019

# Osnova cvičení

- › opakování Hive & SQL
- › import dat do tabulky Hive
- › SQL dotazy nad velkými daty

# Opakování Hive & SQL

- › **tabulka = adresář** (ne soubor)
- › typy tabulek:
  - interní: vlastník adresáře = Hive, umístěno v Hive warehouse
  - externí: vlastník adresáře = někdo jiný, umístěno kdekoliv
- › partition u tabulky:
  - fyzické rozdělení dat do podadresářů
  - rychlejší vyhledání (podmínka WHERE v dotazu)
- › které příkazy DML Hive umí a které ne?
  - **insert** ano (doporučen insert select)
  - **update** a **delete** standardně ne
- › formáty a komprese
  - ORC, Avro, Parquet, text (CSV)
  - Gzip, Snappy, Zlib

# Import dat do tabulky Hive

Viz příklad na přednášce.

Syntaxe CREATE TABLE:

```
CREATE [EXTERNAL] TABLE tablename (  
    field type, ...  
)  
[PARTITIONED BY (field type)] ← pole pro  
[ROW FORMAT] partitioning se  
    uvede jen zde  
    DELIMITED FIELDS TERMINATED BY string  
    LINES TERMINATED BY string  
STORED AS format  
LOCATION hdfs_path  
tblproperties ("key"="value", ...)  
    např. "skip.header.line.count"="1"
```

# Import dat do tabulky Hive

Viz příklad na přednášce.

Syntaxe CREATE TABLE:

```
INSERT [OVERWRITE | INTO] TABLE tablename
[PARTITION (field)]
SELECT
...
FROM tablename
```

← pole pro partitioning se uvede nahoře i zde

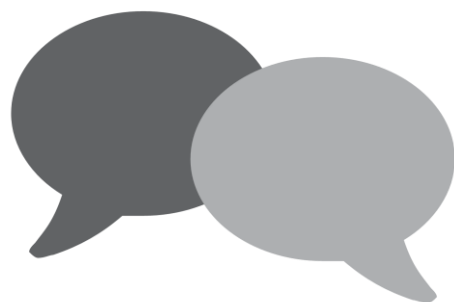
## dynamický partitioning

```
set hive.exec.dynamic.partition=true;
```

```
set hive.exec.dynamic.partition.mode=nonstrict;
```

# Procvičování

- › <https://github.com/stameser/BDT>
- › cviceni/03\_HIVE
- › vzorová řešení: branch solutions



## **Diskuze**

# Díky za pozornost

**PROFINIT**

NÁSKOK DÍKY ZNALOSTEM

Profinit EU, s.r.o.  
Tychonova 2, 160 00 Praha 6



Telefon  
+ 420 224 316 016



Web  
[www.profinit.eu](http://www.profinit.eu)



LinkedIn  
[linkedin.com/company/profinit](https://linkedin.com/company/profinit)



Twitter  
[twitter.com/Profinit\\_EU](https://twitter.com/Profinit_EU)